NCI HINTS Sample Design and Weighting Plan

Lou Rizzo, Westat

May 5, 2003

The sample design is a list-assisted RDD sample from all telephone exchanges in the U.S., with oversampling of exchanges with high numbers of blacks and Hispanics. This will result in a nationally representative sample of households. During the household screener, one adult was sampled within each household and recruited for the extended interview.

The list-assisted RDD method is a random sample of telephone numbers from all 'working banks' in U.S. telephone exchanges (see for example Tucker, Casady, and Lepkowski 1993). A working bank is a set of 100 telephone numbers (e.g., telephone numbers with area code 301 and first five digits 294-44) with at least one listed residential number[1].

1. Within Household Sampling

Our plan was to sample one adult within each sampled household, with each adult having an equal chance of selection. One approach is the last birthday method which is described for example in Binson, Canchola, and Catania (2000). We ask the respondent how many adults are in the household, and then ask which adult has had the most recent birthday. That adult becomes the selected adult.

Our proposed plan was designed to be as unintrusive as possible while still giving each adult an equal chance of selection. The steps are as follows:

- Confirm that the respondent is an adult. Ask the respondent how many adults are in the household. The respondent answers N=1, 2, 3, ...
- The CATI system accesses a preselected random number RAND1 for the sampled household.
  - If RAND1 is less than or equal to 1/N, then the respondent is selected. No further sampling steps are necessary (note that if N is 1, the respondent is automatically selected).
  - If RAND1 is greater than 1/N, and N=2, then the respondent is informed that his/her housemate has been selected. No further sampling steps are necessary.
  - If RAND1 is greater than 1/N, and N>2, then adult sampling continues.
    - The respondent is asked if he/she knows the birthdays of his/her housemates.
      - If the respondent says yes, then the respondent is asked to identify the housemate with the most recent birthday (excluding themselves). No further sampling steps are necessary.
      - If the respondent says no, then the respondent is asked to give the first names of his/her housemates (or first initials). CATI will sort these names in alphabetical order. A preselected random number RAND2 will

[1] Note that all numbers whether listed as residential or not are part of the sampling frame, as long as they are in working banks.

be accessed to choose one adult. (For example if there are two housemates named Jim and Mary, then CATI sorts the names as Jim followed by Mary. If RAND2 is less than or equal to 0.5, then Jim is selected. If RAND2 is greater than 0.5, then Mary is selected.)

This sampling plan has the virtue of minimizing the number of questions asked of the respondent. If there are one or two adults in the household, then only the first question about number of adults is necessary. If there are N=3 or more adults, then there is a 1/N chance that the respondent is selected precluding further questions. Otherwise, the birthday method should be sufficient in most cases to complete adult sampling. The intrusive question about first names and first initials is only a last resort which is not asked often.

The preselection of respondent or not allows us to avoid the possibility of self-selection which may be occurring with the standard birthday method.

2. Oversampling Minorities

Part of the protocol for the NCI HINTS study is a requirement to achieve high precision for the subdomain of blacks and Hispanics. To achieve this we oversampled from telephone exchanges which have a higher percentage of blacks and Hispanics. Our subcontractor Genesys has estimates of the percentage of minorities for each active telephone exchange in the U.S. We've studied a number of different possible stratifications for past RDD surveys, and the best we believe is to assign to a high minority stratum all exchanges with an estimated 15 percent or more blacks and Hispanics, with the complement set of exchanges becoming the low-minority stratum. If we oversample this high-minority stratum we can increase the expected number of black and Hispanic households. We want to oversample at a high enough rate to give us maximum percentage standard errors no higher than 0.015 for both black and Hispanic domain estimates.

In computing standard errors for this allocation and others we need to account for the design effect produced by the within-household sampling. Using March 1998 Current Population Survey data we can estimate that 31.9 percent of households have one adult, 53.8 percent of households have two adults, and 14.3 percent of households have three or more adults, with a mean number of 3.9 adults in the three or more adult households. We assign a household weighting factor of 1 for the one-adult households, a household weighting factor of 2 for the two-adult households (to account for sampling of one of two adults), and a larger weighting factor for the remaining households. A rough estimate for the design effect incurred for these differential weighting factors is $(1+CV^2)$, where CV is the coefficient of variation in the weighting factors (see Kish 1992). Our estimate from the CPS data is 1.22. In terms of standard errors, this is an increase of 10.5 percent from the simple random sampling result. The effective sample sizes below incorporate this Kish factor of 1.22.

We can achieve 1.6% standard errors for blacks and Hispanics with 8,000 extended interviews with the following sample design. We proposed to reach 14,000 households, and expected 80 percent response at the screener level from the recruited households, resulting in 11,200 completed screeners. These households and screeners were allocated to the high minority and low minority in a 66:41 proportion (oversampling the high minority stratum, as the relative share of the two strata in the population is 51.2:48.8). This was done by oversampling the high-minority exchanges at a 1.5428 rate (i.e., each high-minority exchange telephone number has a probability of selection 1.5428 times higher than the low-minority exchange telephone numbers).

Table 1 presents estimates as to how the expected households and completed screeners were allocated to black, Hispanic, and other domains within the two sampling strata given the oversampling of the high-minority stratum at the 1.5428 rate.

Our plan was then to subsample other race/ethnicity families in the high minority stratum at a 64.8% rate (the reciprocal of 1.5428). This equates the final sampling rate for other race/ethnicity families within the high-minority exchange stratum to that of other race/ethnicity families within the low-minority exchange stratum, improving efficiency for this group (as we do not need 'extra' households within this domain). The expected attempted interviews in Table 1 reflect the expected sample sizes at this point. The final column of Table 1 is the expected final interviews, which are the expected attempted interviews multiplied by 0.85, the expected extended interview response rate.

Table 1. Stratum-domain percentages and sample sizes for proposed sample design[2].

Stratum	Domain	Portion of pop'n	Portion of stratum	Expected attempted screeners	Expected completed screeners	Family subsmpg rate	Expected attempted interviews	Expected final interviews
High minrty	Black	10.47%	20.47%	1,760	1,408	100.0%	1,408	1,152
High minrty	Hispanic	10.72%	20.95%	1,801	1,441	100.0%	1,441	1,179
High minrty	Other	29.98%	58.58%	5,038	4,030	64.8%	2,612	2,137
High minrty	All	51.17%	100.0%	8,600	6,880		5,461	4,469
Low minrty	Black	1.35%	2.77%	150	120	100.0%	120	98
Low minrty	Hispanic	1.64%	3.36%	181	145	100.0%	145	119
Low minrty	Other	45.84%	93.87%	5,069	4,055	100.0%	4,055	3,319
Low minrty	All	48.83%	100.0%	5,400	4,320		4,320	3,536
All	All			14,000	11,200		9,781	8,005

Using the formulas for the variances of a stratified random sample with the Kish factor for the design effect for within-household sampling, we have the following results for this sample design:

Table 2. Standard errors and effective sample sizes for black and Hispanic domains.

Domain	Expected completed interviews	Standard error	Effective sample size
Black domain	1,250	1.58%	1,002
Hispanic domain	1,298	1.55%	1,037
All adults	8,005	0.62%	6,472

[2] This table is an updated version of the table given in January 10, 2002 version of this report, and is based on final percentages of minorities within the two exchange strata as calculated at the time of RDD sampling.

The standard errors are for population percentages of 0.5. The effective sample size is the sample size for a simple random sample which would achieve the same precision for population percentages of 0.5.

3. RDD Sample Sizes

Based on our sample sizes needs and our projections of eligibility rates and response rates at the initiation point of the project, we made the following projections. We assumed that roughly 43% of the telephone numbers are residential in the working banks. We assumed a screener response rate of 80% and an extended interview response rate of 85%. Table 3 presents our expected breakdown for the telephone sample, with the total telephone sample size required as 32,560.

Table 3. Overall RDD telephone sample sizes.

Total Telephone Sample	32,560
Total Households in Sample (43 percent residential rate)	14,000
Total Expected Completed Screeners (80 percent response rate)	11,200
Total Attempted Interviews (after subsampling)	9,425
Total Completed Interviews (85 percent response rate)	8,000

This RDD sample size was divided into two waves, with 16,280 numbers in Wave 1 and 16,280 numbers in Wave 2, and Wave 1 was fielded. In the course of fielding Wave 1, we found that our residency rates and response rates were much lower than expected. In response to this, we added 50% to the Wave 2 sample, and added a further reserve sample of 12,210 numbers as well.

Table 4. RDD wave and reserve totals.

	Telephone sample size
Wave 1	16,280
Wave 2	24,420
Reserve	12,210
Total	52,910

We also applied a two-phase stratification approach to the reserve sample, based on whether or not the telephone numbers had mailable addresses associated with them or not. This is introduced in Brick, et al (2002). The nonmailable numbers were subselected at an 80% rate, i.e., 20% of these numbers were discarded from the sample. We weight the remaining nonmailable numbers at a rate of 1.25 to offset this subsampling. This leads to an increase in variance from the differential weighting, but the nonmailable numbers are much less productive, so that the tradeoff leads to better variance properties.

Table 5. Two-phase stratification of the reserve sample numbers.

	Total numbers in original sample	Percent	Total numbers in final sample	Percent	Weighting factor
Mailable	8,730	71.50%	8,730	75.82%	1.00
Nonmailable	3,480	28.50%	2,784	24.18%	1.25
Total reserve	12,210	100.00%	11,514	100.00%	

4. Sample Weights and Variance Estimation

Every sampled adult who completed a questionnaire in HINTS will receive a sampling weight and a set of replicate sampling weights. These sampling weights should be used in aggregating any survey questionnaire answers for the purpose of computing nationally representative estimates.

The sampling weight consists of three major components. The first component is the respondent's base weight. This base weight is the reciprocal of the probability that the respondent had of being sampled. Section 5 discusses the computation of base weights. The second part of the sampling weight is an adjustment for nonresponse. There are several points at which cooperation needs to be gained: the household needs to be successfully reached and the screener successfully completed, and the sampled respondent within the household needs to be successfully recruited to complete the extended (HINTS) interview. Both a screener nonresponse adjustment and an extended interview nonresponse adjustment will be computed. The computation of the screener nonresponse adjustment is complicated by the fact that many residential households are never reached even after a considerable number of calls, and are never completely confirmed as residential or nonresidential. These telephone numbers with unknown residential status can be categorized as NM numbers (for which only an answering machine is reached) and NA numbers (for which no contact is made of any kind). Section 6 discusses nonresponse adjustments in detail.

The third part of the sampling weight is a calibration adjustment. The primary purpose of the calibration adjustment is to reduce the sampling variance of estimators through the utilization of reliable auxiliary information (reliable in the sense of having less sampling and nonsampling error than the corresponding HINTS estimates). For example, the total number of male and female adults in the United States is estimable by taking the summation of all (nonresponse-adjusted) base weights of responding adults in the survey by sex. There are other estimates of these same population totals with less sampling and nonsampling error that can be used to calibrate the HINTS estimates (e.g., if HINTS population estimates for males deviate from corresponding estimates from the auxiliary information, the weights of male respondents can be altered to bring HINTS estimates “in line” with the auxiliary information). This process of calibration improves the sampling error of HINTS estimates which are correlated in the population with characteristics represented in the auxiliary information. The auxiliary information used for HINTS will come from the most recent Current Population Survey (probably March

2001), which has much larger sample sizes than HINTS. These calibration adjustments are discussed in Section 8 below.

5. Jackknife Variance Estimation

The sampling weights for each responding adult are sufficient for the computation of statistically sound nationally representative estimators based on HINTS data. It is also necessary to produce statistically valid standard errors for these estimators.

The jackknife technique is compatible with the sample design and weighting procedures for HINTS. The jackknife variance estimation technique takes carefully selected subsets of the data for each "replicate", and for each respondent in the replicate subset determines a sampling weight, as if the replicate subset were in fact the responding sample (this replicate subset is usually almost all of the sample, except for a group of respondents which are "deleted" for that replicate). The resulting weights are called replicate weights.

The jackknife procedure is the standard operating procedure for variance estimation at Westat for surveys such as HINTS. The Westat software package, WesVarPC, can be used to calculate these variance estimators. It can be obtained from the Internet by accessing the WesVarPC site:

http://www.westat.com/wesvar/

We will also retain on the output data files the necessary implicit and explicit stratification variables necessary to use linearization software packages such as SUDAAN.

The jackknife variance estimator will be computed in the following way for HINTS. A set of $R=50$ replicate weights was assigned to each responding adult. Suppose **P** is a percentage of adults in the U.S. population having a particular characteristic (e.g., answering one of the HINTS questions in a particular way). A nationally representative estimator p can be computed by aggregating the adult sampling weights of all responding adults with this characteristic (e.g., all responding adults in the survey answering the survey question in a particular way). A jackknife variance estimator of the sampling variance of p can be computed in two steps:

1. Recompute estimators $p(r)$, $r=1,\dots,R$, by aggregating the replicate sampling weights corresponding to replicate r for all responding adults with the characteristic.

2. Compute the jackknife variance estimator

$$v(p) = \frac{R-1}{R}\sum_{r=1}^{R}(p(r)-p)^2$$

The replicate weights will be computed by systematically deleting a portion of the original sample, and recomputing the sampling weights as if the remaining sample (without the deleted portion) were the actual sample. These deleted sample units should be first-stage sampling units, which in HINTS are telephone households. The remainder of the sample with the deleted portion removed is called the 'replicate subset', and it should mirror the full sample design, as if it were a reduced version of the original sample. The HINTS RDD sample is a stratified sample, so each replicate subset has a sample from each stratum, with that sample reduced by the deleted portion (i.e., each stratum has a piece deleted from it of roughly equal size, guaranteeing that the replicate subset interpenetrates every stratum to an equal degree).

For the purposes of jackknife variance estimation each sample telephone number will be assigned to one of 50 replicate "deletion" groups $D(r)$, $r=1,...,50$. This will be done in such a way that $1/50$ of each sampling stratum will be assigned to each group $D(r)$ (i.e., the deletion groups include parts of each sampling stratum). Each replicate sample is the full sample minus the deletion group (i.e., it is roughly 49/50 of the original sample).

The replicate sampling weights will be generated in a series of steps that parallel the steps computing the full sample sampling weights. The replicate base weight for each sampled household or adult and each replicate is either equal to $R/(R-1)$ times the full sample base weight (if the household is contained in the replicate subset) or equal to 0 (if the household is not contained in the replicate subset, but instead is contained in the "deleted" set for that replicate). See Section 5 for further details on computation of the replicate base weights.

Nonresponse and poststratification adjustments will then be computed for each set of replicate base weights, using the replicate base weights in the computation of nonresponse and poststratification adjustments in place of the original base weights. These calculations generate a set of replicate nonresponse and poststratification adjustments for each responding adult. The final replicate weights are products of the replicate base weights, nonresponse adjustments, and poststratification adjustments. Sections 6 and 7 discuss in detail the computation of the final replicate weights.

6. Base Weights

Base weights will be assigned to both sampled households and sampled adults within households. The base weight for the respondent is the product of four factors:

- the reciprocal of the telephone number's probability of being selected in the RDD sample (i.e., the sampling rate);
- the reciprocal of the conditional probability of the respondent being selected among the adults in the household given that the household was selected (which is equal to the number of adults in the household);
- An extra factor equal to 2 if the household has more than one regular, residential telephone number, which accounts for the doubled chance of selection of the household[3].
- An extra factor of 1.25 if the household was a nonmailable number in the reserve sample (to offset the 80% subsampling of these numbers).

The base weight will be indicated below as w_i (i indicating the particular sampled adult).

[3] Note that cellular numbers, numbers devoted to businesses run from the household, and numbers dedicated to fax or modem usage are not considered. There are a small number of households with more than two regular, residential telephone numbers, but this number is small. We simplify the questionnaire by only asking about one or more than one, and use that information in the computation of the base weight.

Standard errors will be computed for HINTS estimates through the use of the jackknife technique, as discussed in Section 5. A total of 50 replicate base weights will be computed for each sample unit[4]. Suppose we write as *A* the set of all sampled adults in the study. Any given survey estimate can be written as

$$Y = \sum_{i \in A} w_i y_i$$

where y_i is the value of a particular survey characteristic for sampled adult *i*, and w_i is the full-sample base weight. The *r*-th replicate estimate for *Y* can be written as:

$$Y(r) = \sum_{i \in A} w_i(r) y_i \qquad \text{with } w_i(r) = \begin{cases} \frac{R}{R-1} w_i & i \in A(r) \\ 0 & i \in D(r) \end{cases}$$

The set *A(r)* is the replicate set corresponding to replicate *r*, and the set *D(r)* is the deleted set corresponding to replicate *r* (see Section 5 for a description of the sampling procedure to select the deleted sets). The union of *A(r)* and *D(r)* for each replicate *r* is the full sample set *A*.

7. Nonresponse Adjustment and Response Rates

Nonresponse is generally encountered to some degree in every survey. The first and most obvious effect of nonresponse is to reduce the effective sample size, which increases the sampling variance. In addition, if there are systematic differences between the respondents and the nonrespondents, there also will be a bias of unknown size and direction. This bias is generally adjusted for in the case of unit nonrespondents (nonrespondents who refuse to answer any part of the questionnaire) with the use of a weighting adjustment term multiplied to the base weights of sample respondents. Item nonresponse (nonresponse to specific questions only) is generally adjusted for through the use of imputation. This section discusses weighting adjustments for unit nonresponse, and calculations of response rates.

The most widely accepted paradigm for unit nonresponse weighting adjustment is the quasi-randomization approach (Oh and Scheuren, 1983). In this approach, nonresponse cells are defined based on those measured characteristics of the sample members that are known to be related to response propensity. For example, if it is known that males respond at a lower rate than females, then sex should be one characteristic used in generating nonresponse cells.

Under this approach, sample units are assigned to a response cell, based on a set of defined characteristics. The weighting adjustment for the sample unit is the reciprocal of the estimated response rate for the cell. Any set of response cells must be based on characteristics which are known for all sample units, responding and nonresponding. Thus questionnaire answers on the survey cannot be used in the development of response cells, because these characteristics are only known for the responding sample units.

---

[4] The total of 50 was chosen from among a number of acceptable alternatives. Generally a large number is necessary for stable variance estimates (e.g., greater than 10), but a number much greater than for example 100 generates sample files that are too large in size (because of large number of replicate weight fields).

Under the quasi-randomization paradigm, we model nonresponse as a "sample" from the population of students in that cell. If this model is in fact valid, then the use of the quasi-randomization weighting adjustment eliminates any nonresponse bias (see for example Little and Rubin (1987), Chapter 4).

7.1 Unit Nonresponse in HINTS

There will be two types of unit nonresponse in HINTS: screener nonresponse and extended interview nonresponse. Screener nonresponse occurs when a household is reached, but the screener interview is not completed. We also need to include in any screener nonresponse calculations any households for which we never reached a person, either because we only reached an answering machine (these are called NM numbers), or only got a ring with no answer (these are called NA numbers), with every call made to the telephone number. Since we don't know if an answering machine or ring no answer corresponds to a residential household, the number of lost residential numbers among the NA and NM numbers needs to be estimated (see Section 7.3 below).

To adjust for screener nonresponse, each completed screener received a screener nonresponse adjustment equal to the reciprocal of the estimated response rate in its screener nonresponse cell. For a discussion of the screener nonresponse cells and adjustments see Sections 7.2 and 7.3 respectively.

Extended interview nonresponse occurs when the screener interview is completed successfully, yielding a sampled adult in the household with identifying information for this adult, and the number of adults in the household, but the sampled adult does not complete the extended interview.

To adjust for interview nonresponse, each completed extended interview receives an interview nonresponse adjustment equal to the reciprocal of the weighted interview response rate in its interview nonresponse cell. (Completed extended interviews also receive a screener nonresponse adjustment.) The methodology for selecting extended interview nonresponse cells and computing extended interview nonresponse adjustments is discussed in Section 7.4. Section 7.5 discusses the computation of replicate nonresponse adjustments.

7.2 Nonresponse Cells for Screener Nonresponse Adjustments

Nonresponse cells will be generated using cross-classifications based on selected sociodemographic characteristics estimated for each telephone exchange (by our vendor Genesys), and mailable status (whether or not an address was available for the telephone number to send them a letter).

The estimated exchange percentages from Genesys will be used to assign each exchange to cells based on the following characteristics:

- Four cells based on geography (Census region): Northeast, South, Midwest, West;

- Three cells (with roughly equal populations)[5] by percent college graduates (exchanges with lowest percentage, next lowest percentage, and highest percentage);
- Three cells (with roughly equal populations) by median income;
- Three cells by percent blacks and Hispanics (two within the high-minority stratum, and the undivided low-minority stratum).

We judge that these characteristics may be both related to response propensity and correlated to item response to HINTS questionnaire items, so that these cells will lead to effective nonresponse adjustments.

The other characteristic for generating cells is the mailable, non-mailable status indicating whether or not a published address is available for the telephone number. These addresses will be used to mail advanced letters about the study and follow-up letters for households who have not responded. We have found in previous surveys that response propensity may differ by this characteristic (telephone households with known addresses which have received mailed information respond at a higher rate than those without known addresses).

Cross-classifications of these sociodemographic classes and the mailout status gives a potential total of 216 cells (though some of the cells may be empty). We will collapse these cells to attain a minimum cell size of 10 sample units and a maximum cell adjustment of 3.0, using our in-house COLL_ADJ software.

6.3 Screener Nonresponse Adjustments

In general, nonresponse adjustments within nonresponse cells are the reciprocals of the weighted response rates within the cell, where the respondents and nonrespondents are weighted by their (adjusted) base weight. In this case, the household base weights are unknown for screener nonrespondents, since components of the base weight depend on whether the household has one or more residential telephone numbers. For this reason, the nonresponse adjustment is set equal to the reciprocal of the unweighted screener response rate for each cell.

In principle, the unweighted screener response rate is equal to the total number of cooperating households (eligible or not) divided by the total number of residential numbers in the sample. The latter value is not completely known, because of NM and NA numbers. Let *AMNA(a)* and *PNA(a)* be the counts of NM and NA numbers in cell *a*. We will estimate the number of residential numbers among the NM numbers by computing the overall eligibility rate *EM* among working numbers with known eligibility status, and by computing the overall eligibility rate *EA* among all numbers with known eligibility status (working and non-working).

With these two estimated eligibility rates applied to the NM and NA numbers, the nonresponse adjustment for cell *a* will be computed as follows:

$$HNRA(a) = \frac{C(a) + I(a) + REF(a) + O(a) + \left(AMNA(a) * EM\right) + \left(PNA(a) * EA\right)}{C(a) + I(a)}$$

[5] The breakpoints will be the 1/3 and 2/3 percentiles over all frame exchanges, which will be calculated when the frame is constructed.

where $C(a)$ is the number of completed screeners, $I(a)$ is the number of households found ineligible for the study, $REF(a)$ is the number of eligible screeners who refused to participate, and $O(a)$ are other residential numbers (e.g., numbers which were found to be residential, but for which a screening interview could not be completed for reasons other than refusals).

We will also compute a study screener response rate. Writing *C, I, REF, O, AMNA*, and *PNA* as the total number of completed screeners, ineligibles, eligible screeners who refused to participate, other residentials, answering machine NA's, and pure NA's respectively, and defining *EM* and *EA* as above, we will compute the screener response rate *SCRNR* as

$$SCRNR = \frac{C + I}{C + I + REF + O + (AMNA * EM) + (PNA * EA)}$$

Note that this screener response rate is algebraically equivalent to

$$SCRNR = \frac{C}{C + \{ER * [REF + O + (AMNA * EM) + (PNA * EA)]\}}$$

with $$ER = \frac{C}{C + I}$$

The second form of *SCRNR* though algebraically more complicated is conceptually more transparent. The response rate is the completes divided by the completes plus the estimated eligible numbers among the remaining residential numbers (refusals and NA's). We estimate the eligibles among the estimated residential numbers $REF+O+(AMNA*EM)+(PNA*EA)$ by imputing the eligibility rate from the 'known eligibility status' numbers: the completes and ineligibles. *SCRNR* is fully within the guidelines of AAPOR standards regarding valid response rates[6].

7.4 Extended Interview Response Cells

There is more information available about extended interview nonrespondents as compared to screener nonrespondents. This extra information comes from the completed screener (a case is not designated as an extended interview nonrespondent unless the screener is successfully completed). In this section, a screener is defined as completed if the key items for sampling an adult and assigning a base weight to the household are answered: the number of adults in the household and the presence of multiple telephone numbers. Note that only if the screener is complete are we able to compute the base weight w_i (see Section 5).

Extended interview nonresponse cells will be generated using cross-classifications of the following characteristics of the sampled adult and household:

1. Sex of sampled adult.
2. Size of household: number of adults in household (1, 2, or more than 2).
3. Census region (4 cells)
4. Telephone number in high, medium, or low minority exchange (3 cells).

[6] Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. Available on AAPOR (American Association for Public Opinion Research) website www.aapor.org.

5. Telephone number in high, medium, or low college educated exchange (3 cells);
6. Telephone number in high, medium, or low median income exchange (3 cells).

The first two characteristics on the list are derived directly from the screener questionnaire for the. The remaining four characteristics are derived from the telephone exchange, and are identical to those utilized in constructing screener nonresponse cells (see Section 7.2). Nonresponse cells will be constructed by collapsing the initial cells to meet the criteria that the cell sample size should be no smaller than 10 and the nonresponse adjustment should be no bigger than 3.0. This will be done using Westat's in-house software COLL_ADJ.

Weighted nonresponse adjustments will be computed for each extended interview cell *b* as follows:

$$ENRA(b) = \frac{\sum_{i \in SA(b)} w_i HNRA(a)}{\sum_{i \in SRA(b)} w_i HNRA(a)}$$

where w_i is the base weight for sampled adult *i*, *SA(b)* is the set of all sampled adults (in cooperative screeners) in interview response cell *b*, *SRA(b)* is the set of all sampled adults in cell *b* completing an extended interview (i.e., the extended interview respondents), and $HNRA(a)$ is the screener nonresponse adjustment for the screener nonresponse cell *a* containing household *i*. The denominator of *ENRA(b)* is an unbiased estimator (adjusted for screener nonresponse[7]) of the total number of adults in the nonresponse cell who would answer an extended interview if contacted (the "population respondents"), the numerator of *ENRA(b)* is an unbiased estimator of the total number of adults in the nonresponse cell (also adjusted for screener nonresponse), and *ENRA(b)* is an approximately unbiased estimator of the response rate which would be obtained in cell *b* if the entire U.S. population were contacted for the study.

We will also compute a weighted extended interview response rate, for reporting purposes. Write SA as the set of all sampled adults from completed screeners and SRA as the set of all sampled adults completing an extended interview. The weighted extended interview response rate is computed as follows:

$$EXTINR = \frac{\sum_{i \in SRA} w_i HNRA(a)}{\sum_{i \in SA} w_i HNRA(a)}$$

[7] Under full response, the sum of the base weights is an unbiased estimator. With the presence of nonresponse, there will be nonresponse bias from any differences between the responding and nonresponding households. This nonresponse bias is reduced in magnitude by the screener nonresponse adjustments. We can't expect these adjustments to eliminate all bias, so the claim of "unbiasedness" of these totals needs to receive this caveat.

7.5. Replicate Nonresponse Adjustments

Nonresponse adjustments are themselves random variables, and contribute a variance component to the overall sampling variance. This variance component is represented in the final jackknife estimator by replicating the computation of nonresponse adjustments (by replacing the original base weights by the replicate base weights, and repeating the computations described in Sections 7.2, 7.3, and 7.4).

The screener nonresponse adjustments are the reciprocals of unweighted screener response rates. Replicate screener response rates were computed for each screener response cell *a* and each replicate *r* by removing the deleted set (of telephone numbers) corresponding to each replicate *r* and recomputing the response rate. In other words, we recomputed response rates for each replicate set as if it were the original RDD sample of telephone numbers.

Define *RS(a,r)* as the count of confirmed residential numbers in screener response cell *a* which are in replicate set *r*. (An alternative definition of *RS(a,r)* is the count of confirmed residential numbers in screener response cell *a* after the deleted set corresponding to replicate *r* has been removed from the RDD sample.) Define *AMNA(a,r), PNA(a,r), EM(r), EA(r), C(a,r), I(a,r), REF(a,r),* and *O(a,r)* similarly (see Section 7.3). Then we can define a replicate nonresponse adjustment as follows:

$$HNRA(a,r) = \frac{C(a,r) + I(a,r) + REF(a,r) + O(a,r) + \left(AMNA(a,r) * EM(r)\right) + \left(PNA(a,r) * EA(r)\right)}{C(a,r) + I(a,r)}$$

The computation of interview nonresponse adjustments will also be replicated. The replicate interview nonresponse adjustment for interview nonresponse cell *b* and replicate *r* is computed as follows:

$$ENRA(b,r) = \frac{\sum_{i \in SA(b)} w_i(r) HNRA(a,r)}{\sum_{i \in SRA(b)} w_i(r) HNRA(a,r)}$$

The two nonresponse adjustments (for screener nonresponse and extended interview nonresponse) are appended to the base weight for the subject (adult):

$$SBW_i = w_i \; HNRA(a) \; ENRA(b)$$

The summation of these nonresponse-adjusted subject base weights over all responding subjects is a nonresponse-adjusted unbiased estimator of the total number of adults in the U.S. population. The corresponding replicate weights are as follows (for each replicate *r*):

$$SBW_i(r) = w_i(r) \; HNRA(a,r) \; ENRA(b,r)$$

8. Calibration Adjustments

The purpose of calibration is to reduce the sampling variance of estimators through the use of reliable auxiliary information. One recent source for this theory is Deville and Sarndal (1992). In the ideal case, this auxiliary information usually takes the form of known population totals for particular characteristics (called control totals). However, calibration also reduces the sampling variance of estimators if the auxiliary information has sampling errors, as long as these sampling errors are significantly smaller than those of the survey itself.

Calibration reduces sampling errors particularly for estimators of characteristics that are highly correlated to the calibration variables in the population. The extreme case of this would be the calibration variables themselves. The survey estimates of the control totals would have considerably higher sampling errors than the “calibrated” estimates of the control totals, which would be the control totals themselves. The estimator of any characteristic that is correlated to any calibration variable will share partially in this reduction of sampling variance, though not fully. Only estimators of characteristics that are completely uncorrelated to the calibration variables will show no improvement in sampling error. Deville and Sarndal (1992) provide a rigorous discussion of these results.

8.1 Control Totals from the Current Population Survey

The Current Population Survey (CPS) of the U.S. Bureau of the Census has much larger sample sizes than those of HINTS. The CPS estimates of any U.S. population totals have lower sampling error than the corresponding HINTS estimates, making calibration of the survey weights to CPS control totals beneficial. The CPS estimates are available via the internet: we will utilize the most current estimates available on the Census website.

Any potential calibration variable needs to be on the CPS public use file, and to be well-correlated to important HINTS questionnaire item outcomes (i.e., we want CPS-available characteristics which tend to have differing mean values for HINTS questionnaire item outcomes). We believe the following CPS characteristics will correlate well with HINTS questionnaire items:

g1. Sex
- 1) Male
- 2) Female

g2. Race/ethnicity
- 1) Hispanic
- 2) NonHispanic black
- 3) NonHispanic white or other.

g3. Age
- 1) 18 to 34 years old
- 2) 35 to 49 years old
- 3) 50 to 64 years old
- 4) 65 years old or older.

g4. Educational Level
- 1) Less than high school diploma
- 2) High school diploma only
- 3) High school diploma, some college

4) Bachelor's degree or higher

Our plan is to generate 96 potential poststratification cells based on cross-classifications of these characteristics, and tabulate control totals for these cells from the most current CPS data. We can write these control totals as c_g, g=1,…,96. The poststratification adjustments are computed by adjusting the weights SBW_i of adult respondents by a constant factor so that the final set of weights add to the control total. I.e., for each poststratification cell *PS(g)*, we compute CA_g equal to:

$$CA_g = \frac{c_g}{\sum_{i \in PS(g)} SBW_i} \quad .$$

We will check each cell to make sure that the sample size is at least 20 and the adjustment is not too much larger than the other adjustments. If any cell has too small a sample size or has too large an adjustment, we will collapse it with other cells. Westat's in-house software COLL_ADJ does this adjustment process with input from a statistician, if it is necessary.

Replicate versions of the *CA* adjustments will also be computed for each replicate *r*. The replicate $CA_g(r)$ adjustments are computed using the same formula, but with the replicate $SBW_i(r)$ weights replacing the full sample SBW_i weights. These replicate versions of the *CA* adjustments are indicated as $CA_g(r)$.

The final subject weights W_i are assigned to each subject *i* with a completed interview. These weights are equal to the base weights for subject *i* multiplied to nonresponse adjustments and calibration adjustments, and can be written as:

$$W_i = SBW_i \, CA_g$$

The replicate weight for subject *i* and replicate *r* is as follows:

$$W_i(r) = SBW_i(r) \, CA_g(r)$$

References

Binson, D., Canchola, J. A., and Catania, J. A. (2000). 'Random selection in a national telephone survey: a comparison of the Kish, next-birthday, and last-birthday methods', *Journal of Official Statistics, 16*, 53-59.

Brick, J. M., Judkins, D., Montaquila, J., and Morganstein, D. (2002). 'Two-phase list-assisted RDD sampling', in *Journal of Official Statistics, 18*, 203-215.

Deville, J. C., and Sarndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association, 87*, 376-382.

Little, R., and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. New York, NY: John Wiley & Sons.

Kish, L. (1992). 'Weighting for unequal π_i', *Journal of Official Statistics 8*, 183-200.

Oh, H., and Scheuren, F. (1983). Weighting adjustments for unit response. In W.G. Madow, I. Olkin, and D. B. Rubin (Eds.), *Incomplete Data in Sampling Surveys, Vol. II: Theory and Annotated Bibliography*. New York: Academic Press.

Tucker, C., Casady, R., and Lepkowski, J. (1993). 'A hierarchy of list-assisted stratified telephone sample design options'. Paper presented at the Annual Conference of the American Association for Public Opinion Research, St. Charles, IL.